HW1. 2022-18599 이기범

#1.

(a) $l_i(\theta) = \frac{1}{2}(x_i^T\theta - Y_i)^2$. Let $x_i^T = [x_{i1}, x_{i2}, \cdots, x_{ip}]$

$$= \frac{1}{2}\left(\sum_{j=1}^{p} x_{ij}\cdot\theta_j - Y_i\right)^2.$$

$$\frac{\partial}{\partial\theta_j} l_i(\theta) = \left(\sum_{j=1}^{p} x_{ij}\cdot\theta_j - Y_i\right)\cdot x_{ij} = (x_i^T\theta - Y_i)\cdot x_{ij}$$

$\therefore \nabla_\theta l_i(\theta) = (x_i^T\theta - Y_i)\cdot x_i$

(b) $\mathcal{L}(\theta) = \frac{1}{2}\|X\theta - Y\|^2 = \frac{1}{2}\sum_{i=1}^{N} 2l_i(\theta) = \sum_{i=1}^{N} l_i(\theta)$

$$\therefore \nabla_\theta \mathcal{L}(\theta) = \sum_{i=1}^{N}\nabla_\theta l_i(\theta) = \sum_{i=1}^{N}(x_i^T\theta - Y_i)x_i = \sum_{i=1}^{N}(x_i^T\theta x_i - Y_i x_i)$$

$$= X^TX\theta - X^TY. \quad \left(\because x_i^T\theta x_i = \begin{pmatrix} x_{i1}^2\theta_1 \\ \vdots \\ x_{ip}^2\theta_p \end{pmatrix} = x_i^Tx_i\theta\right)$$

#2. $\theta^{k+1} = \theta^k - \alpha f'(\theta^k) = \theta^k - \alpha\theta^k = (1-\alpha)\theta^k$.

if $\alpha > 2$, $|\theta^{k+1}| = |1-\alpha|\cdot|\theta^k|$.

By induction, $|\theta^{k+1}| = |1-\alpha|^{k+1}\cdot|\theta^0|$. $|1-\alpha| > 1$.

For $\forall N$, let $k = \left\lceil \log_{|1-\alpha|}\frac{N}{|\theta^0|}\right\rceil$. then

$|\theta^k| = |1-\alpha|^k|\theta^0| \geq \frac{N}{|\theta^0|}\cdot|\theta^0| = N$. hence θ^k diverges.

#3. $\nabla f(\theta^k) = X^T(X\theta^k - Y)$ due to (1b).

$\therefore \theta^{k+1} = \theta^k - \alpha X^T(X\theta^k - Y) = (1 - \alpha X^TX)\theta^k + \alpha X^TY$.

Say $\theta^* = (X^TX)^{-1}X^TY$.

$$\theta^{k+1} - \theta^* = (I - \alpha X^TX)\theta^k + \alpha X^TY - \theta^*$$

$$= (I - \alpha X^TX)(\theta^k - \theta^*) + \alpha X^TY - \alpha X^TX\theta^*$$

$$= (I - \alpha X^T X)(\theta^k - \theta^*) + \cancel{\alpha X^T Y - \alpha X^T X (X^T X)^{-1} X^T Y}$$

$$\|\theta^{k+1} - \theta^*\| = \det(I - \alpha X^T X)\,\|\theta^k - \theta^*\| .$$

$$\det(I - \alpha X^T X) = \prod (\text{eigen values of } \alpha X^T X)$$

$$= \prod (1 - \alpha \cdot \lambda_i)$$

λ_i : eigen value of $X^T X$ ($\lambda_i > 0$)

$$\alpha > \frac{2}{\rho(X^T X)} \geq \frac{2}{\lambda_i}, \text{ so } 1 - \alpha \cdot \lambda_i < 1 - 2 = -1 .$$

Hence, $M = \left|\det(I - \alpha X^T X)\right| > 1$.

Because $\|\theta^{k+1} - \theta^*\| = M \|\theta^k - \theta^*\| = M^{k+1} \|\theta^0 - \theta^*\|$,

If $\theta^0 \neq \theta^*$, $\|\theta^k - \theta^*\|$ diverges, which means that θ^k diverges as well because θ^* is fixed / finite.

$\mu(\{\theta^*\}) = 0$, so it diverges a.e. $\square$